CODES FOR KEY GENERATION IN QUANTUM CRYPTOGRAPHY

BERTHOLD-GEORG ENGLERT
Department of Physics, National University of Singapore, Singapore 117542
phyebg@nus.edu.sg

FANG-WEI FU*
Temasek Laboratories, National University of Singapore, Singapore 117508
tslfufw@nus.edu.sg

HARALD NIEDERREITER and CHAOPING XING
Department of Mathematics, National University of Singapore, Singapore 117543
nied@math.nus.edu.sg, matxcp@nus.edu.sg

Received 12 April 2005

As an alternative to the usual key generation by two-way communication in schemes for quantum cryptography, we consider codes for key generation by one-way communication. We study codes that could be applied to the raw key sequences that are ideally obtained in recently proposed scenarios for quantum key distribution, which can be regarded as communication through symmetric four-letter channels.

Keywords: Error correcting codes, linear codes, quantum key distribution

1. Introduction

In a recently proposed protocol for quantum key distribution,[1,2] Alice sends uncorrelated qubits through a quantum channel to Bob. Under ideal circumstances, the channel is noiseless, and then the situation is as follows.

Alice prepares each qubit in one of four states — labeled A, B, C, and D, respectively, and chosen at random — and Bob detects each qubit in one of four states that are labeled correspondingly. The set-up has the peculiar feature that Bob *never* obtains the letter that specifies the state prepared by Alice. Rather, he always gets one of the other three letters, whereby the laws of quantum physics ensure that the outcome is truly random, and each possibility occurs equally likely.

These physical laws also prevent any third party, eavesdropper Eve, from acquiring information about Alice's or Bob's letters. Therefore, they can exploit the

*On leave from the Department of Mathematics, Nankai University, Tianjin 300071, P. R. China

correlations between their letters to generate a private cryptographic key, which they can then use for the secure encryption of a message.

The key generation is a crucial step. Two different procedures are described in Refs. 1 and 2, with respective efficiencies of $\frac{1}{3}$ and $\frac{2}{5}$ key bits per letter. Both procedures rely on *two-way* communication between Alice and Bob. By contrast, it is our objective here to study codes for the key generation by *one-way* communication.

After the exchange of many qubits through the quantum channel, Alice and Bob have random sequences of the four letters, such that corresponding letters are never the same, while each of the twelve pairs of different letters occurs one-twelfth of the time, with no correlations between the pairs. Alice sends a code word to Bob by telling him, through a public channel, the positions at which the letters appear in her sequence — such as "3rd letter, then 14th, 15th, 92nd, and 65th" for a particular five-letter word. Bob forms the received word from his corresponding letters, and then decodes.

The public communication does not leak any useful information to Eve. Thus, if Alice chooses a random sequence of code words, each word being equally likely, as she will do, Eve knows nothing about Alice's words. She also knows nothing about Bob's decoded words, provided that Bob's decoding procedure does not favor some words at the expense of others. Accordingly, the sequence of words constitutes a privately shared key for secure classical communication between Alice and Bob.

There is a nonzero probability that Bob's received word is consistent with two or more words that Alice could have sent, so that the decoding will not be completely error-free. A good, practical code must, therefore, represent a compromise between (i) having not too many code words, (ii) an acceptable error rate, and (iii) a reasonable efficiency. Arguably the best compromise we report in Section 8 is code (3) of Example 1. It has 1024 words, an error rate of 0.6%, and an efficiency of $\frac{1}{4}$ key bits per letter.

As there is no fundamental reason why the key generation by one-way communication should be substantially less efficient than that by two-way communication, one expects that more efficient codes can be found. Therefore, the work reported here should be regarded as a first step, not as the final word on the matter.

It is worth mentioning that there is a very similar problem for the three-letter channel of Renes's "trine" scheme.[1] Further, the standard BB84 protocol[3] has a four-letter channel with quite different properties, for which codes for one-way key generation are not known. The same remark applies to the six-letter generalization[4] of BB84. In short, there is a whole class of coding problems that deserve attention.

2. Probability Distributions

The quantum protocols of Alice and Bob involve two random variables X and Y taking values in $\{A, B, C, D\}$. We have the following corresponding probability distributions with $x, y \in \{A, B, C, D\}$. The joint probability distribution of X and

Y is given by

$$\Pr\{X = x, Y = y\} = \begin{cases} 0 & \text{if } y = x\,, \\ \dfrac{1}{12} & \text{if } y \neq x\,. \end{cases} \tag{1}$$

Accordingly, the marginal probability distributions of X and Y are

$$\Pr\{X = x\} = \frac{1}{4}\,, \qquad \Pr\{Y = y\} = \frac{1}{4}\,, \tag{2}$$

and the conditional probability distribution of Y with respect to X is

$$\Pr\{Y = y \mid X = x\} = \begin{cases} 0 & \text{if } y = x\,, \\ \dfrac{1}{3} & \text{if } y \neq x\,. \end{cases} \tag{3}$$

Now we compute the information-theoretic quantities entropy, conditional entropy, and mutual information[a] of the random variables X and Y. The entropy of Y is

$$H(Y) = -\sum_{y\in\{A,B,C,D\}} \Pr\{Y = y\} \log_2 \Pr\{Y = y\} = 2\,, \tag{4}$$

and for the conditional entropy of Y with respect to X we find

$$H(Y|X) = -\sum_{\substack{x,\,y\,\in\,\{A,\,B,\,C,\,D\} \\ y\,\neq\,x}} \Pr\{X = x, Y = y\} \log_2 \Pr\{Y = y|X = x\} = \log_2 3\,, \tag{5}$$

and we obtain

$$\begin{aligned} I(X;Y) &= \sum_{\substack{x,\,y\,\in\,\{A,\,B,\,C,\,D\} \\ y\,\neq\,x}} \Pr\{X = x, Y = y\} \log_2 \frac{\Pr\{X = x, Y = y\}}{\Pr\{X = x\}\Pr\{Y = y\}} \\ &= H(Y) - H(Y|X) = \log_2 \frac{4}{3} \end{aligned} \tag{6}$$

for the mutual information of X and Y.

3. Discrete Memoryless Channel

The information transmission from Alice to Bob can be described in information theory by a discrete memoryless channel.[a] This channel is characterized by the conditional probability distribution[a] of Y with respect to X,

$$Q(y|x) = \Pr\{Y = y \mid X = x\} = \begin{cases} 0 & \text{if } y = x\,, \\ \dfrac{1}{3} & \text{if } y \neq x\,, \end{cases} \tag{7}$$

[a]For the definitions of these and other information-theoretic quantities see Ref. 5, for example.

where $x,\ y \in \{A, B, C, D\}$. The *channel capacity*[a] is defined by

$$c = \max_{P_X} I(X;Y) = \max_{P} I(P;Q)\,, \tag{8}$$

where

$$I(P;Q) = \sum_{x,y} P(x)Q(y|x) \log_2 \frac{Q(y|x)}{\sum_{x'} P(x')Q(y|x')} \tag{9}$$

and the maximum is taken over all probability distributions P on $\{A, B, C, D\}$.

The channel defined by (7) is a symmetric channel (see Theorem 8.2.1 on p. 190 in Ref. 5). Hence, the capacity is attained by the uniform distribution on $\{A, B, C, D\}$, so that

$$c = \log_2 \frac{4}{3} \doteq 0.4150\,. \tag{10}$$

For any positive integer n, the nth extension of this discrete memoryless channel has the conditional probability distribution

$$Q^n(\mathbf{y}|\mathbf{x}) = \begin{cases} 0 & \text{if } y_i = x_i \text{ for some } i, \\ \dfrac{1}{3^n} & \text{if } y_i \neq x_i \text{ for all } i, \end{cases} \tag{11}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \{A, B, C, D\}^n$ are n-letter words.

4. Codes for the Specific Channel

In this section, we discuss the design of codes and decoding methods for the specific channel introduced in Section 3.

Let $\mathbf{F}_4$ be the finite field with four elements. It is convenient to let A, B, C, D be represented respectively by the four elements 0, 1, a, b of $\mathbf{F}_4$ since we want to use linear codes for this specific channel. The addition and multiplication tables of $\mathbf{F}_4$ are as follows:

$$\begin{array}{c|cccc} + & 0 & 1 & a & b \\ \hline 0 & 0 & 1 & a & b \\ 1 & 1 & 0 & b & a \\ a & a & b & 0 & 1 \\ b & b & a & 1 & 0 \end{array} \qquad \begin{array}{c|cccc} \times & 0 & 1 & a & b \\ \hline 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & a & b \\ a & 0 & a & b & 1 \\ b & 0 & b & 1 & a \end{array} \tag{12}$$

Let $\mathbf{F}_4^n$ be the n-dimensional vector space over $\mathbf{F}_4$. For two vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{F}_4^n$, $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in \mathbf{F}_4^n$, the *Hamming distance* $d(\mathbf{x}, \mathbf{y})$ between $\mathbf{x}$ and $\mathbf{y}$ is defined as the number of coordinates in which they differ,

$$d(\mathbf{x}, \mathbf{y}) = \bigl|\{i : x_i \neq y_i\}\bigr|\,. \tag{13}$$

The *Hamming weight* $w(\mathbf{x})$ is the number of nonzero coordinates in $\mathbf{x}$,

$$w(\mathbf{x}) = \bigl|\{i : x_i \neq 0\}\bigr|\,. \tag{14}$$

A code of length n with $M \geq 2$ codewords is a subset $\mathcal{C}$ of $\mathbf{F}_4^n$,

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}\,, \qquad \mathbf{c}_i \in \mathbf{F}_4^n\,. \tag{15}$$

The *minimum distance* $d(\mathcal{C})$ of the code $\mathcal{C}$ is the minimum Hamming distance between two distinct codewords,

$$d(\mathcal{C}) = \min\{d(\mathbf{x}, \mathbf{y}) : \mathbf{x}, \mathbf{y} \in \mathcal{C},\ \mathbf{x} \neq \mathbf{y}\}\,. \tag{16}$$

We denote by

$$L(\mathbf{c}_i) = \{\mathbf{x} \in \mathbf{F}_4^n :\ d(\mathbf{x}, \mathbf{c}_i) = n\} \quad \text{for } i = 1, 2, \ldots, M \tag{17}$$

the set of n-letter words that could be received if codeword $\mathbf{c}_i$ is sent. It is easy to see that for any $i, j, l \in \{1, 2, \ldots, M\}$ we have

$$\big|L(\mathbf{c}_i)\big| = 3^n\,, \tag{18}$$

$$\big|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\big| \geq 2^n\,, \tag{19}$$

$$\big|L(\mathbf{c}_i) \cap L(\mathbf{c}_j) \cap L(\mathbf{c}_l)\big| \geq 1\,. \tag{20}$$

Note, in particular, the significance of (20): For any three different codewords that could have been sent by Alice, there is at least one word received by Bob that is consistent with all three.

Further, it follows from (11) that

$$Q^n(\mathbf{y}|\mathbf{x}) = \begin{cases} 0 & \text{if } \mathbf{y} \notin L(\mathbf{x})\,, \\ \dfrac{1}{3^n} & \text{if } \mathbf{y} \in L(\mathbf{x})\,. \end{cases} \tag{21}$$

This means that if Alice sends $\mathbf{x} \in \mathbf{F}_4^n$ through this channel to Bob, then Bob receives $\mathbf{y} \in L(\mathbf{x})$ with probability $1/3^n$.

Now we describe a decoding method for the code $\mathcal{C}$ by decoding regions, which exploits the significance of $L(\mathbf{c}_i)$. The M subsets $D_1, D_2, \ldots, D_M$ of $\mathbf{F}_4^n$ are called *decoding regions* for the code $\mathcal{C}$ if they satisfy the following conditions:

$$\begin{aligned} &\text{(i) } D_i \subseteq L(\mathbf{c}_i)\,, \quad i = 1, 2, \ldots, M\,; \\ &\text{(ii) } D_i \cap D_j = \emptyset\,, \quad i \neq j\,; \\ &\text{(iii) } \bigcup_{i=1}^{M} D_i = \bigcup_{i=1}^{M} L(\mathbf{c}_i)\,. \end{aligned} \tag{22}$$

The decoding method for the code $\mathcal{C}$ with the decoding regions $D_1, D_2, \ldots, D_M$ is then: decode the received vector $\mathbf{y}$ into $\mathbf{c}_i$ if $\mathbf{y} \in D_i$. In some cases, for simplicity, we can construct the decoding regions in accordance with

$$\begin{aligned} D_1 &= L(\mathbf{c}_1)\,, \\ D_2 &= L(\mathbf{c}_2) \setminus L(\mathbf{c}_1)\,, \\ D_i &= L(\mathbf{c}_i) \setminus \left(\bigcup_{j=1}^{i-1} L(\mathbf{c}_j)\right)\,, \quad i = 2, 3, \ldots, M\,. \end{aligned} \tag{23}$$

This means that we decode the received vector $\mathbf{y} \in \bigcup_{i=1}^{M} L(\mathbf{c}_i)$ into the first $\mathbf{c}_i$ such that $d(\mathbf{y}, \mathbf{c}_i) = n$.

Since the decoding regions of (23) refer to an agreed-upon order of the codewords, the decoding is biased toward the early codewords in the list at the expense of the later ones. Such a bias is avoided by the *maximum likelihood decoding*.[6] It can be described as follows: The received vector $\mathbf{y} \in \bigcup_{i=1}^{M} L(\mathbf{c}_i)$ is decoded into any codeword $\mathbf{c}_i$ such that $Q^n(\mathbf{y}|\mathbf{c}_i)$ is the maximum value of $Q^n(\mathbf{y}|\mathbf{c})$ over all codewords $\mathbf{c} \in \mathcal{C}$. If there is more than one such $\mathbf{c}_i$, we choose one of them at random. It is easy to see that this is equivalent to the following decoding method: The received vector $\mathbf{y}$ is decoded into either one of the codewords $\mathbf{c}_i$ that obey $d(\mathbf{y}, \mathbf{c}_i) = n$, choosing one at random if there are several such $\mathbf{c}_i$s.

5. Decoding Error Probability

In this section, we discuss the decoding error probability and Shannon's Channel Coding Theorem for the specific channel introduced in Section 3. Some criteria for good codes for this channel are given.

For a code $\mathcal{C}$ with decoding regions $D_1, D_2, \ldots, D_M$, the probability e_i of the event that the vector $\mathbf{y}$ received by Bob is not decoded into the codeword $\mathbf{c}_i$ sent by Alice is given by

$$\begin{aligned} e_i &= \Pr\{\mathbf{y} \notin D_i \mid \mathbf{c}_i \text{ is sent}\} = 1 - \Pr\{\mathbf{y} \in D_i \mid \mathbf{c}_i \text{ is sent}\} \\ &= 1 - \frac{|D_i|}{3^n}, \quad i = 1, 2, \ldots, M . \end{aligned} \tag{24}$$

The *average error probability* $\bar{e}$ is the arithmetic mean of the e_is,

$$\begin{aligned} \bar{e} = \frac{1}{M} \sum_{i=1}^{M} e_i &= 1 - \frac{1}{3^n M} \sum_{i=1}^{M} |D_i| \\ &= 1 - \frac{1}{3^n M} \left| \bigcup_{i=1}^{M} D_i \right| \\ &= 1 - \frac{1}{3^n M} \left| \bigcup_{i=1}^{M} L(\mathbf{c}_i) \right| , \end{aligned} \tag{25}$$

and the *maximum error probability* $e_{\max}$ is the largest one of them,

$$e_{\max} = \max_{1 \le i \le M} e_i. \tag{26}$$

Obviously, $\bar{e} \le e_{\max}$. Note that

$$\begin{aligned} \bar{e} = 0 &\iff e_{\max} = 0 \\ &\iff e_i = 0 \text{ for all } i \\ &\iff L(\mathbf{c}_i) \cap L(\mathbf{c}_j) = \emptyset , \quad i \ne j . \end{aligned} \tag{27}$$

Hence, it follows from (19) that

$$e_{\max} \ge \bar{e} > 0 . \tag{28}$$

In particular, if the decoding regions $D_1, D_2, \ldots, D_M$ are given by (23), then

$$\begin{aligned} e_1 &= 1 - \frac{1}{3^n}\big|L(\mathbf{c}_1)\big| = 0\,, \\ e_i &= 1 - \frac{1}{3^n}\left|L(\mathbf{c}_i) \setminus \left(\textstyle\bigcup_{j=1}^{i-1} L(\mathbf{c}_j)\right)\right| \\ &= \frac{1}{3^n}\left[\big|L(\mathbf{c}_i)\big| - \left|L(\mathbf{c}_i) \setminus \left(\textstyle\bigcup_{j=1}^{i-1} L(\mathbf{c}_j)\right)\right|\right] \\ &= \frac{1}{3^n}\left|L(\mathbf{c}_i) \cap \left(\textstyle\bigcup_{j=1}^{i-1} L(\mathbf{c}_j)\right)\right| \\ &= \frac{1}{3^n}\left|\textstyle\bigcup_{j=1}^{i-1}\left(L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right)\right|\,, \quad i = 2, 3, \ldots, M\,. \end{aligned} \tag{29}$$

The decoding error probability of a code is one of its most important performance characteristics. In this connection, we recall Shannon's Channel Coding Theorem (see Refs. 5 and 6).

Theorem 1: (Shannon's Channel Coding Theorem) For any $0 < \varepsilon < 1$ and $0 < R < \log_2(4/3)$, there exists for sufficiently large n a code $\mathcal{C}$ of length n and size $M \doteq 2^{nR}$ such that $e_{\max} \leq \varepsilon$, and so in particular $\bar{e} \leq \varepsilon$.

Remark 1: For fixed n and M, the best we can do is to choose a code

$$\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M\}\,, \qquad \mathbf{c}_i \in \mathbf{F}_4^n\,, \tag{30}$$

such that $\left|\bigcup_{i=1}^{M} L(\mathbf{c}_i)\right|$ is as large as possible, i.e., the average error probability $\bar{e}$ is as small as possible. For fixed n and $\bar{e} \leq \varepsilon$, in view of (25) we have to try to find a code $\mathcal{C}$ with the largest size M such that

$$\frac{1}{M}\left|\bigcup_{i=1}^{M} L(\mathbf{c}_i)\right| \geq 3^n(1-\varepsilon)\,. \tag{31}$$

Note that for certain values of n and ε, such a code $\mathcal{C}$ may not exist.

6. Upper Bounds on the Decoding Error Probability

In this section, we give several upper bounds on the decoding error probability of codes for our specific channel.

Let $\mathcal{C}$ be a code for our specific channel. The *distance distribution* of the code $\mathcal{C}$ is defined by

$$A_s = \frac{1}{M}\big|\{(\mathbf{x}, \mathbf{y}) \in \mathcal{C} \times \mathcal{C} : d(\mathbf{x}, \mathbf{y}) = s\}\big|\,, \quad s = 0, 1, \ldots, n\,. \tag{32}$$

It is easy to see that

$$A_0 = 1\,, \quad \sum_{s=0}^{n} A_s = |\mathcal{C}| = M\,. \tag{33}$$

Theorem 2: Let $\mathcal{C}$ be a code for the specific channel in Section 3. Suppose that the distance distribution of the code $\mathcal{C}$ is given by $A_0, A_1, \ldots, A_n$. Then the average error probability $\bar{e}$ is upper bounded by

$$\bar{e} \leq \frac{1}{2} \sum_{s=1}^{n} A_s \left(\frac{2}{3}\right)^s . \tag{34}$$

Proof: By the definition of $L(\mathbf{c}_i)$ in (17) we know that if $d(\mathbf{c}_i, \mathbf{c}_j) = s$, then

$$\left|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right| = 3^{n-s} 2^s = 3^n \left(\frac{2}{3}\right)^s , \quad s = 1, 2, \ldots, n . \tag{35}$$

Hence,

$$\sum_{1 \leq i < j \leq M} \left|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right| = \frac{3^n M}{2} \sum_{s=1}^{n} A_s \left(\frac{2}{3}\right)^s . \tag{36}$$

By (36) and noting that $|L(\mathbf{c}_i)| = 3^n$, we obtain

$$\begin{aligned} \left|\bigcup_{i=1}^{M} L(\mathbf{c}_i)\right| &\geq \sum_{i=1}^{M} |L(\mathbf{c}_i)| - \sum_{1 \leq i < j \leq M} \left|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right| \\ &= 3^n M - \frac{3^n M}{2} \sum_{s=1}^{n} A_s \left(\frac{2}{3}\right)^s . \end{aligned} \tag{37}$$

Hence, (34) follows from (25) and (37). □

Theorem 3: Let $\mathcal{C}$ be a code for the specific channel in Section 3. If the minimum distance $d(\mathcal{C}) \geq d$, then the average error probability $\bar{e}$ is upper bounded by

$$\bar{e} \leq \frac{M-1}{2} \left(\frac{2}{3}\right)^d < \frac{M}{2} \left(\frac{2}{3}\right)^d . \tag{38}$$

Proof: If the minimum distance $d(\mathcal{C}) \geq d$, then $A_s = 0$ for $1 \leq s \leq d-1$. Hence, by Theorem 2 and (33),

$$\bar{e} \leq \frac{1}{2} \left(\frac{2}{3}\right)^d \sum_{s=d}^{n} A_s = \frac{M-1}{2} \left(\frac{2}{3}\right)^d < \frac{M}{2} \left(\frac{2}{3}\right)^d . \tag{39}$$

This completes the proof. □

Theorem 4: Let $\mathcal{C}$ be a code for the specific channel in Section 3. If the minimum distance $d(\mathcal{C}) \geq d$ and the decoding regions $D_1, D_2, \ldots, D_M$ are given by (23), then the maximum error probability $e_{\max}$ is upper bounded by

$$e_{\max} \leq (M-1) \left(\frac{2}{3}\right)^d < M \left(\frac{2}{3}\right)^d . \tag{40}$$

Proof: If the minimum distance $d(\mathcal{C}) \geq d$, then for $i \neq j$,

$$\left|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right| = 3^{n-d(\mathbf{c}_i,\mathbf{c}_j)} 2^{d(\mathbf{c}_i,\mathbf{c}_j)} = 3^n \left(\frac{2}{3}\right)^{d(\mathbf{c}_i,\mathbf{c}_j)} \leq 3^n \left(\frac{2}{3}\right)^d . \tag{41}$$

It follows from (41) that for $i = 2, 3, \ldots, M$,

$$\begin{aligned} \left|\bigcup_{j=1}^{i-1} \big(L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\big)\right| &\leq \sum_{j=1}^{i-1} \left|L(\mathbf{c}_i) \cap L(\mathbf{c}_j)\right| \leq (i-1) 3^n \left(\frac{2}{3}\right)^d \\ &\leq (M-1) 3^n \left(\frac{2}{3}\right)^d . \end{aligned} \tag{42}$$

Hence, by (29) and (42),

$$e_1 = 0\,, \quad e_i \leq (M-1)\left(\frac{2}{3}\right)^d , \quad i = 2, 3, \ldots, M\,. \tag{43}$$

Therefore, $e_{\max} \leq (M-1)(2/3)^d$. □

7. Linear Codes

In this section, we discuss the design of linear codes and the decoding error probability of linear codes for our specific channel. Some criteria for good linear codes are given.

First, we recall some basic concepts for linear codes. A code $\mathcal{C}$ over $\mathbf{F}_4$ is called a linear $[n, k]$ code over $\mathbf{F}_4$ if $\mathcal{C}$ is a k-dimensional subspace of $\mathbf{F}_4^n$. Note that for a linear $[n, k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ the number of codewords is $M = 4^k$. Furthermore, we call the linear code $\mathcal{C}$ a linear $[n, k, d]$ code if the minimum distance of $\mathcal{C}$ is at least d. Let A_i be the number of codewords in $\mathcal{C}$ of Hamming weight i. The sequence of numbers $A_0, A_1, \ldots, A_n$ is called the *weight distribution* of $\mathcal{C}$. It is well known in coding theory[7] that for a linear code $\mathcal{C}$, the distance distribution of $\mathcal{C}$ is equal to the weight distribution of $\mathcal{C}$ and the minimum distance of $\mathcal{C}$ is equal to the minimum Hamming weight of nonzero codewords. For fixed n and k, let $d_4(n, k)$ be the maximal minimum distance of a linear $[n, k]$ code over $\mathbf{F}_4$. Tables of lower and upper bounds on $d_4(n, k)$ are available in Ref. 8.

We denote by

$$\begin{aligned} \mathcal{A} &= \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{F}_4^n : x_i \neq 0 \text{ for all } i\} \\ &= \{\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{F}_4^n : w(\mathbf{x}) = n\} \end{aligned} \tag{44}$$

the set of vectors with maximal Hamming weight, that is: the set of words that do not have the letter A. For a linear $[n, k]$ code $\mathcal{C}$ over $\mathbf{F}_4$, by (17), it is easy to check that

$$\bigcup_{\mathbf{c} \in \mathcal{C}} L(\mathbf{c}) = \mathcal{A} + \mathcal{C}\,. \tag{45}$$

Hence, by (25), the average error probability $\bar{e}$ can be rewritten as

$$\bar{e} = 1 - \frac{1}{3^n 4^k} |\mathcal{A} + \mathcal{C}| \,. \tag{46}$$

Note that $\mathcal{A} + \mathcal{C}$ is a union of some cosets of $\mathcal{C}$,

$$\mathcal{A} + \mathcal{C} = \bigcup_{j=1}^{\alpha} (\mathbf{a}_j + \mathcal{C}) \,, \quad \mathbf{a}_j \in \mathcal{A} \,, \tag{47}$$

where $\mathbf{a}_1 + \mathcal{C}, \mathbf{a}_2 + \mathcal{C}, \ldots, \mathbf{a}_\alpha + \mathcal{C}$ are some different cosets of $\mathcal{C}$. This implies that $|\mathcal{A} + \mathcal{C}| = \alpha 4^k$. Therefore, the average error probability $\bar{e}$ is also given by

$$\bar{e} = 1 - \frac{\alpha}{3^n} \,. \tag{48}$$

Remark 2: For fixed n and k, the best we can do is to choose a linear $[n, k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ such that α is as large as possible. Note that even if $\mathcal{C}$ is optimal in this sense, the average error probability may not be small. For fixed n and $\bar{e} \le \varepsilon$, we have to try to find a linear $[n, k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ with the largest dimension k such that

$$\alpha \ge 3^n (1 - \varepsilon) \,. \tag{49}$$

Note that for certain values of n and ε, such a linear code $\mathcal{C}$ may not exist.

Remark 3: It is known from Ref. 9 (see Problem 11 on p. 114) that for our specific channel the codes in Shannon's Channel Coding Theorem can be replaced by linear codes over $\mathbf{F}_4$, that is, for any $0 < \varepsilon < 1$ and

$$\frac{k}{n} < \frac{1}{2} \log_2(4/3) \doteq 0.2075 \tag{50}$$

there will exist, for sufficiently large n, a linear $[n, k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ such that

$$\alpha \ge 3^n (1 - \varepsilon) \,. \tag{51}$$

This is equivalent to the fact that $\bar{e} \le \varepsilon$.

In general, for a linear $[n, k, d]$ code $\mathcal{C}$ over $\mathbf{F}_4$, by Theorem 3 we have

$$\bar{e} \le \frac{4^k - 1}{2} \left(\frac{2}{3}\right)^d < \frac{4^k}{2} \left(\frac{2}{3}\right)^d \,, \tag{52}$$

and if the decoding regions $D_1, D_2, \ldots, D_{4^k}$ are given by (23), then by Theorem 4,

$$e_{\max} \le (4^k - 1) \left(\frac{2}{3}\right)^d < 4^k \left(\frac{2}{3}\right)^d \,. \tag{53}$$

Furthermore, if the weight distribution $\{A_s\}_{s=0}^n$ of $\mathcal{C}$ is known, then by Theorem 2,

$$\bar{e} \le \frac{1}{2} \sum_{s=1}^{n} A_s \left(\frac{2}{3}\right)^s \,. \tag{54}$$

By using the Gilbert-Varshamov quasi-random construction of linear codes,[7] one can construct, for sufficiently large n, a linear $[n,k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ with size

$$M = 4^k \doteq 4^{n[1-H_4(d/n)]} \tag{55}$$

such that the minimum distance $d(\mathcal{C}) \geq d$, where

$$H_4(x) = x\log_4 3 - x\log_4 x - (1-x)\log_4(1-x)\,, \quad 0 \leq x \leq \frac{3}{4}\,. \tag{56}$$

It follows from (53) that the maximum error probability $e_{\max}$ is upper bounded by

$$e_{\max} < 4^k \left(\frac{2}{3}\right)^d \doteq 4^{n[1-H_4(d/n)+(d/n)\log_4(2/3)]}\,. \tag{57}$$

The function of d/n in the exponent is such that

$$1 - H_4(x) + x\log_4\frac{2}{3} < 0 \Longleftrightarrow x > \beta\,, \tag{58}$$

where $\beta \doteq 0.4627$ is the unique solution of $1 - H_4(x) + x\log_4(2/3) = 0$. This means that, for sufficiently large n, one can construct a linear $[n,k]$ code $\mathcal{C}$ over $\mathbf{F}_4$ with the rate

$$\frac{k}{n} \doteq 1 - H_4(\beta) \doteq 0.1353 \tag{59}$$

such that the maximum error probability $e_{\max}$ is arbitrarily small.

8. Some Examples

In this section we give some examples of linear codes for the specific channel in Section 3 to illustrate our results. These linear codes are listed in Brouwer's tables[8] of presently best known quaternary linear codes. The value of R is obtained from $M = 2^{nR}$ in Theorem 1 and corresponds to the efficiency mentioned in Section 1.

In Table 1 we give the parameters of various codes from Ref. 8. They illustrate a general observation, namely that there is a trade-off between the simplicity of the code (short length n of the words, small number M of them) on one side and the performance of the code (large efficiency R, small average error $\bar{e}$). In addition to the codes of Table 1, we mention the following four codes of moderate length and reasonably good performance.

Example 1: Here we list explicitly known linear codes of moderate lengths.

(1) A code with $n = 28\,,\ k = 4\,,\ d = 20\,, M = 256$ for which $R = \frac{2}{7} \doteq 0.2857$ and $\bar{e} < \frac{4^4}{2}(2/3)^{20} \doteq 0.03849$ is the upper bound of (52).
As is known from Ref. 8, there exists a linear $[28,4,20]$ code over $\mathbf{F}_4$ with the weight distribution $A_{20} = 189\,,\ A_{24} = 63\,,\ A_{28} = 3$, so that (54) gives the bound $\bar{e} \leq 0.03038$.

Table 1. Examples of linear codes for the channel specified by the conditional probability distribution (7). For each code we give the number n of letters in each codeword, the dimension k of the subspace of $\mathbf{F}_4^n$, the minimum Hamming distance d, the total number M of codewords, the efficiency R, and in the last column an upper bound of (52) on the average error probability $\bar{e}$, rounded to four significant digits.
The first group on the left are six codes of length $n = 100$ with consecutive values of the dimension k. For $k \geq 16$, the upper bound on $\bar{e}$ is greater than 1, and so it is not meaningful. The second group on the left are two codes of large lengths which demonstrate that the error probability can be made very small. The group on the right are 14 codes with lengths decreasing from 50 to 10.

n	k	d	M	R	$\bar{e} \leq$	n	k	d	M	R	$\bar{e} \leq$
100	10	62	4^{10}	0.2	6.337×10^{-6}	50	5	35	1024	0.2	3.516×10^{-4}
100	11	60	4^{11}	0.22	5.704×10^{-5}	50	6	33	4096	0.24	3.165×10^{-3}
100	12	58	4^{12}	0.24	5.133×10^{-4}	48	6	32	4096	0.25	4.747×10^{-3}
100	13	56	4^{13}	0.26	4.620×10^{-3}	48	5	33	1024	0.208	7.912×10^{-4}
100	14	55	4^{14}	0.28	0.02772	47	6	31	4096	0.255	7.120×10^{-3}
100	15	52	4^{15}	0.30	0.3742	46	5	32	1024	0.217	1.187×10^{-3}
						45	5	31	1024	0.222	1.780×10^{-3}
						43	5	30	1024	0.233	2.670×10^{-3}
						42	5	29	1024	0.238	4.005×10^{-3}
200	20	109	4^{20}	0.2	3.517×10^{-8}	41	5	28	1024	0.244	6.008×10^{-3}
250	25	136	4^{25}	0.2	6.340×10^{-10}	40	4	28	256	0.2	1.502×10^{-3}
						30	3	22	64	0.2	4.277×10^{-3}
						20	2	16	16	0.2	0.01218
						10	1	10	4	0.2	0.02601

(2) A code with $n = 31\,,\ k = 4\,,\ d = 22\,,\ M = 256$ for which $R = \frac{8}{31} \doteq 0.2581$ and $\bar{e} < \frac{4^4}{2}(2/3)^{22} \doteq 0.01711$ is the upper bound of (52).
As is known from Ref. 8, there exists a linear $[31, 4, 22]$ code over $\mathbf{F}_4$ with the weight distribution $A_{22} = 141\,,\ A_{24} = 87\,,\ A_{28} = 24\,,\ A_{30} = 3$, so that (54) gives the bound $\bar{e} \leq 0.01216$.

(3) A code with $n = 40\,,\ k = 5\,,\ d = 28\,,\ M = 1024$ for which $R = \frac{1}{4} = 0.25$ and $\bar{e} < \frac{4^5}{2}(2/3)^{28} \doteq 0.006008$ is the upper bound of (52).
As is known from Ref. 8, this optimal linear code is a quasi-cyclic code. The generator matrix can be represented as $G = [G_0, G_1, G_2, G_3, G_4, G_5, G_6, G_7]$ where G_i for $0 \leq i \leq 7$ are 5×5 circulant matrices. The first row of G is given by $[10000 \quad 10120 \quad 11020 \quad 11230 \quad 12220 \quad 13130 \quad 13210 \quad 11312]$, where we identify 2 with $a \in \mathbf{F}_4$ and 3 with $b \in \mathbf{F}_4$.

(4) The shortened code of the previous example: $n = 39\,,\ k = 4\,,\ d = 28\,,\ M = 256$ for which $R = \frac{8}{39} \doteq 0.2051$ and $\bar{e} < \frac{4^4}{2}(2/3)^{28} \doteq 0.001502$.

Remark 4: For codes of small size, one can calculate the exact values of e_i, $\bar{e}$, and $e_{\max}$ by using (24)–(29). The decoding method is also computationally feasible. For codes of large size, for example $M = 4^{20}$, decoding will become an enormous computational task.

Remark 5: By using nonlinear codes, it may be possible to achieve better results on the decoding error probability, but we have not tried to search for good

quaternary nonlinear codes in the literature.

Acknowledgments

This research is supported in part by the DSTA research grant R-394-000-011-422 and in part by ICITI research grant R-144-000-109-112. The work of Fang-Wei Fu is also supported by the National Natural Science Foundation of China (Grant No. 60172060), the Trans-Century Training Program Foundation for the Talents by the Education Ministry of China, and the Foundation for University Key Teacher by the Education Ministry of China.

References

1. J. M. Renes, "Spherical-code key-distribution protocols for qubits," Phys. Rev. A **70**, 052314 (2004).
2. B.-G. Englert, D. Kaszlikowski, H. K. Ng, W. K. Chua, J. Řeháček, and J. Anders, "Highly Efficient Quantum Key Distribution With Minimal State Tomography," e-print quant-ph/0412089.
3. C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *IEEE Conference on Computers, Systems, and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
4. D. Bruß and C. Macchiavello, "Optimal Eavesdropping in Cryptography with Three-Dimensional Quantum States," Phys. Rev. Lett. **88**, 127901 (2002).
5. T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
6. R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.
7. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*. Amsterdam, The Netherlands: North-Holland, 1977.
8. A. E. Brouwer, "Bounds on the minimum distance of linear codes," available at `http://www.win.tue.nl/~aeb/voorlincod.html`
9. I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press, 1981.